ARE211, Fall 2009

NPP1: THU, NOV 5, 2009 PRINTED: AUGUST 27, 2009 (LEC# 21)

Contents

Preliminary draft only: please check for final version

6. NONLINEAR PROGRAMMING PROBLEMS AND THE KARUSH KUHN TUCKER CONDITIONS

Key points:

(1) The canonical form of the nonlinear programming problem (NPP)

$$\text{maximize } f(\mathbf{x}) \text{ subject to } \mathbf{g}(\mathbf{x}) \le \mathbf{b},$$

(2) Conditions for existence of a solution the NPP:

Theorem: If $f : A \to \mathbb{R}$ is continuous and *strictly quasi-concave*, and A is compact, nonempty and convex, then f attains a unique maximum on A.

(3) Conditions relating a local maximum on the constraint set to a unique solution to the NPP:

Theorem: If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and strictly quasi-concave and $g^j : \mathbb{R}^n \to \mathbb{R}$ is *quasi-convex* for each j, then if f attains a local maximum on $A = \{\mathbf{x} \in \mathbb{R}^n : \forall j, g^j(\mathbf{x}) \le b_j\}$, this max is the unique global maximum of f on A.

(4) The KKT conditions in Mantra format:

(Except for a bizarre exception) a necessary condition for $\mathbf{x}$ to solve a constrained maximization problem is that the gradient vector of the objective function at $\mathbf{x}$ belongs to the nonnegative cone defined by the gradient vectors of the constraints that are satisfied with equality at $\mathbf{x}$.

(5) The KKT conditions in math format:

If $\bar{\mathbf{x}}$ solves the maximization problem *and the constraint qualification holds* at $\bar{\mathbf{x}}$ then there exists a vector $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^m_+$ such that

$$\triangledown f(\bar{\mathbf{x}})^T = \bar{\boldsymbol{\lambda}}^T J\mathbf{g}(\bar{\mathbf{x}})$$

Moreover, $\bar{\boldsymbol{\lambda}}$ has the property that $\bar{\lambda}_j = 0$, for each j such that $g^j(\bar{\mathbf{x}}) < b_j$.

(6) The distinction between a constraint satisfied with equality and a binding constraint.

(7) The role of the constraint qualification (CQ):

  (a) it ensures that the linearized version of the constraint set is, locally, a good approximation of the true nonlinear constraint set.

  (b) the CQ will be satisfied at $\mathbf{x}$ if the gradients of the constraints that are satisfied with equality at $\mathbf{x}$ form a linear independent set

(8) Why the KKT conditions are necessary: three special cases.

The general nonlinear programming problem (NPP) is the following:

$$\text{maximize } f(\mathbf{x}) \text{ subject to } \mathbf{g}(\mathbf{x}) \leq \mathbf{b},$$

$$\text{where } f : \mathbb{R}^n \to \mathbb{R} \text{ and } \mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$$

Terminology:

- f is called the *objective function*;
- $\mathbf{g}$ is a vector of m *constraint functions*, and, of course, $\mathbf{b} \in \mathbb{R}^m$. That is, the individual constraints are stacked together to form a vector-valued function.
- The set of $\mathbf{x}$ such that $\mathbf{g}(\mathbf{x}) \leq \mathbf{b}$ is called the *feasible set* or *constraint set* for the problem.

For the remainder of the course, unless otherwise notified, we will assume that both the objective function and the constraint functions are continuously differentiable. Indeed, we will in fact assume that they are as many times continuously differentiable as we could ever need.

Emphasize that this setup is *completely* general, i.e., covers every problem you are ever likely to encounter.

- can handle constraints of the form $\mathbf{g}(\mathbf{x}) \geq \mathbf{b}$;
- can handle constraints of the form $\mathbf{x} \geq 0$;
- can even handle constraints of the form $\mathbf{g}(\mathbf{x}) = \mathbf{b}$.

For example, given $u : \mathbb{R}^2 \to \mathbb{R}$, consider the problem

$$\text{maximize } u(\mathbf{x}) \text{ subject to } \mathbf{p} \cdot \mathbf{x} \leq y, \mathbf{x} \geq 0;$$

What's g: in this case, g is a linear function, defined as follows:

$$\begin{aligned} g^0(\mathbf{x}) &= \mathbf{p} \cdot \mathbf{x} \\ g^1(\mathbf{x}) &= -x_1 \\ g^2(\mathbf{x}) &= -x_2 \end{aligned}$$

so that the problem can be written as

$$\text{maximize } u(\mathbf{x}) \text{ subject to } G\mathbf{x} \leq \mathbf{b},$$

$$\text{where } G = \begin{bmatrix} p_1 & p_2 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} y \\ 0 \\ 0 \end{pmatrix}$$

While there are many advantages to having a single, general version of the KT conditions, the generality comes at a (small) cost. When it comes to actually *computing* the solution to an NPP problem (as opposed to just understanding what's going on), it's very convenient to treat equality constraints *differently* from inequality constraints. I explain what you need to do on page 17. *You should make sure to refer to this discussion before you start on the NPP problem set.*

6.1. **Existence and Uniqueness**

The first step is: under what conditions does the NPP have a solution at all? Under what conditions will the solution be unique? We'll answer this question by considering when a function defined on an arbitrary set A attains a maximum and when this maximum is unique. Then we'll apply these results to the NPP, letting A denote the constraint set for the problem. Recall the following theorem, also known as the *extremum value theorem.*

Theorem (Weierstrass): If $f : A \to \mathbb{R}$ is continuous and A is compact, nonempty, then f attains a maximum on A.

- relationship between A and the constraint sets of the NPP: A is the intersection of the lower contour sets defined by the NPP, i.e., $g_j(\cdot) \leq b^j$, for all j.
- role of compact, i.e., closed and bounded, and nonempty; *not* guaranteed in the specification of the NPP.
  - example of why existence of a maximum may fail if A is not closed: try to maximize $f(x) = 1/x$ on the interval $(0, 1]$.

  - example of why existence of a maximum may fail if A is not bounded try to maximize $f(x) = x$ on $\mathbb{R}_+$.
- when you set up the problem, have to make sure the constraint set you end up with is nonempty and compact, otherwise you may not get a solution.

Theorem: If $f : A \to \mathbb{R}$ is continuous and *strictly quasi-concave*, and A is compact, nonempty and convex, then f attains a unique maximum on A.

- example of why uniqueness may fail if f is not quasi-concave: let A be a circle and construct a function f which has a level set that has two points of tangency with the ball.
- example of why uniqueness may fail if A is not convex: let A be a ball with a bite out of it, and let f be a quasi-concave function whose level set touches the edge of the constraint surface, where the surface has been bitten, at two distinct points.

Theorem: If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and strictly quasi-concave and $g^j : \mathbb{R}^n \to \mathbb{R}$ is *quasi-convex* for each j, then if f attains a local maximum on $A = \{\mathbf{x} \in \mathbb{R}^n : \forall j, g^j(\mathbf{x}) \leq b_j\}$, this max is the unique global maximum of f on A.

Proof:

- recall that the *upper* contour sets of a quasi-concave function are convex sets.
- similarly, the *lower* contour sets of a quasi-convex function are convex sets.
- the condition that $g^j(\mathbf{x}) \leq b_j$ is the condition that $\mathbf{x}$ lie in the lower contour set of the j'th constraint identified by b_j.
- finally, the intersection of convex sets is convex.
- now apply previous theorem about uniqueness given conditions on A.

Thus, quasi-convexity of g^j's guarantees convexity of A. However, it doesn't imply COMPACTNESS, hence existence isn't guaranteed.

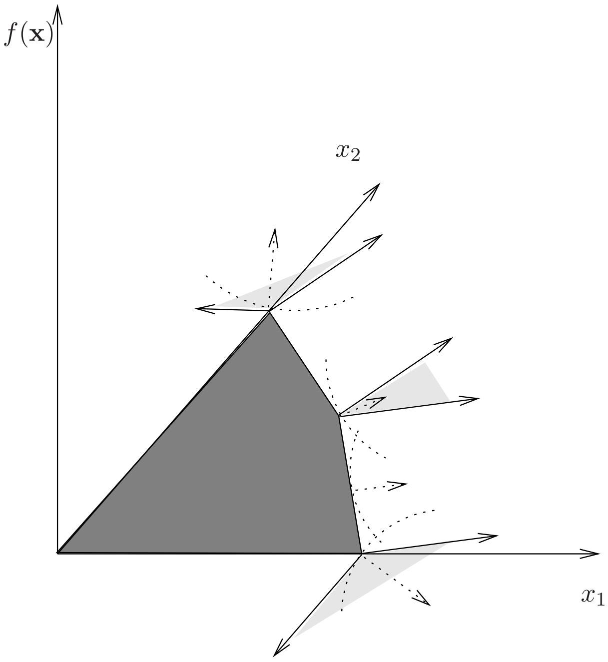

Figure 1. The KKT conditions

Take stock: haven't told you how to solve one of these problems, only when it will have a solution and when that solution will be unique.

6.2. **Necessary conditions for a solution to an NPP**

In this topic, we'll give two theorems, one is necessary and one is sufficient for $\bar{\mathbf{x}}$ to be a maximum. For now, we're focusing on necessary conditions. First, the theorem in words, in Mantra format:

Theorem (Kuhn Tucker necessary conditions) as a Mantra: (Except for a bizarre exception) a necessary condition for $\mathbf{x}$ to solve a constrained maximization problem is that the gradient vector of the objective function at $\mathbf{x}$ belongs to the nonnegative cone defined by the gradient vectors of the constraints that are satisfied with equality at $\mathbf{x}$.

The bizarre exception is called the *constraint qualification.* We'll get to it in a minute.

Now for the formal statement of the theorem.

Theorem (Kuhn Tucker necessary conditions):[1] If $\bar{\mathbf{x}}$ solves the maximization problem *and the constraint qualification holds* at $\bar{\mathbf{x}}$ then there exists a vector $\bar{\boldsymbol{\lambda}} \in \mathbb{R}^m_+$ such that

$$\triangledown f(\bar{\mathbf{x}})^T = \bar{\boldsymbol{\lambda}}^T J\mathbf{g}(\bar{\mathbf{x}})$$

Moreover, $\bar{\boldsymbol{\lambda}}$ has the property that $\bar{\lambda}_j = 0$, for each j such that $g^j(\bar{\mathbf{x}}) < b_j$.

The scalar λ_j is known as the *lagrange multiplier* corresponding to the j'th constraint. Note that $\boldsymbol{\lambda}^T$ denotes the transpose of $\boldsymbol{\lambda}$, i.e., if $\boldsymbol{\lambda}$ is assumed to be a column vector, then $\boldsymbol{\lambda}^T$ will be a row vector. Why didn't I write $J\mathbf{g}(\bar{\mathbf{x}})\boldsymbol{\lambda}$?

- Recall that if $A\mathbf{x} = \mathbf{b}$, then $\mathbf{b}$ is a weighted linear combination of the *columns* of A.
- Imagine if you wrote $\mathbf{x}'A = \mathbf{b}$, then what would $\mathbf{b}$ be? Ans.: it is a weighted linear combination of the *rows* of A.
- That's what we want here: in words, what the theorem says is: if $\bar{\mathbf{x}}$ solves the maximization problem *and the constraint qualification holds* then the gradient of f at $\bar{\mathbf{x}}$ is a nonnegative weighted linear combination of the gradients of the constraints that are satisfied with equality.

In some math-for-economists text-books (e.g., Simon and Blume), the term "constraint satisfied with equality" is (unfortunately) replaced by "binding constraints". This is unfortunate because in practice, economists almost always reserve the term "binding" for a subclass of the constraints satisfied with equality. Specifically, a constraint is said to be *binding* if the maximized value of the objective function increases when the constraint is relaxed a little bit. (A constraint is relaxed by increasing the constant on the right hand side of the constraint inequality.) For example in Fig. 2, both constraints are satisfied with equality, but only g^1 is binding; if you relax constraint g^1 the optimum moves to the north-east; if you relax constraint g^2, the solution is unchanged. Formally,

---

[1] Actually, this is not the Kuhn Tucker theorem but the Karusch Kuhn Tucker theorem. Karusch was a graduate student who was partly responsible for the result. Somehow, his name has been forgotten. A regrettable thing for graduate students

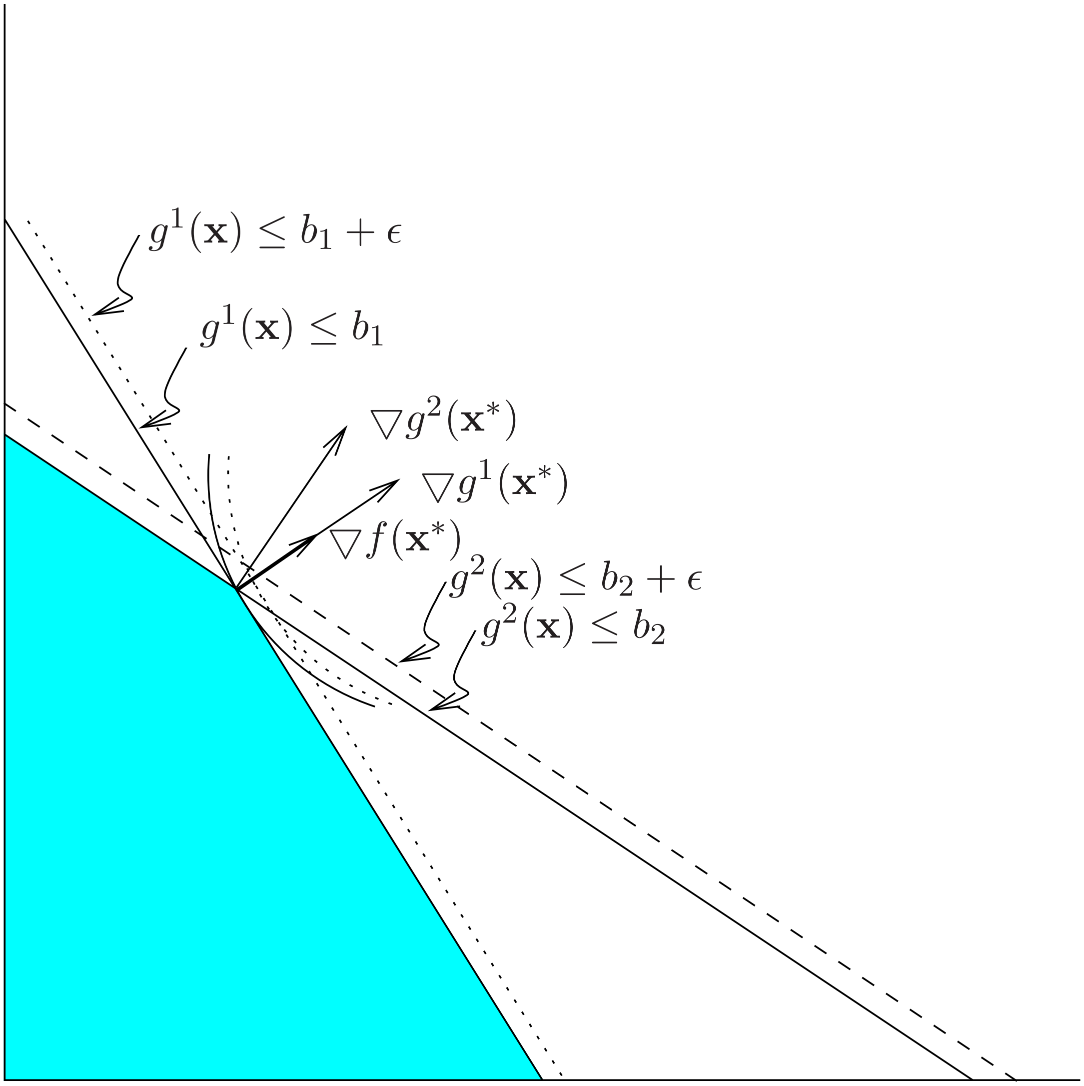

FIGURE 2. constraint g^1 is binding, but g^2 isn't

consider the problem

$$\max f(\mathbf{x}) \text{ s.t. } \mathbf{g}(\mathbf{x}) \leq \mathbf{b}$$

Let $M(\mathbf{b})$ denote the solution to the problem with constraint vector $\mathbf{b}$. We now say that constraint j is *binding* if there exists $\bar{\epsilon} > 0$ such that $\forall \epsilon \in (0, \bar{\epsilon})$, $M(\mathbf{b} + \epsilon e^j) > M(\mathbf{b})$, where e^j denotes the j'th unit vector. In this course we'll always say that a constraint is "satisfied with equality" unless we know that it's actually binding in the sense above.

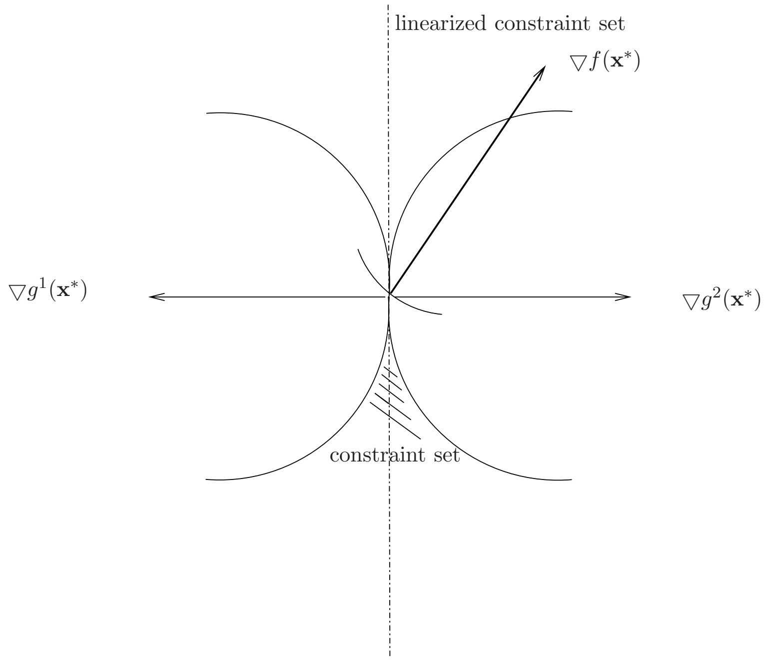

FIGURE 3. Varian's "Hershey Kiss" example: the Constraint Qualification is violated

Economists almost always decide whether a constraint is binding or merely satisfied with equality depending on whether the corresponding lagrange multiplier is positive or zero. This is almost but not quite correct: positivity of the lagrangian is *sufficient* but not necessary for bindingness. For example, consider max $f(x) = x^3$, such that $g(x) = x \leq b = 0$. The solution to this problem is $x^* = 0$. At this point, $f'(x^*) = 0 \times g'(x^*)$, i.e., the lagrangian is zero. But the constraint is certainly binding. We've seen this distinction over and over again in this course. What's going on here is exactly the same as what's going on when we say that negative definiteness of the Hessian is sufficient but not necessary for strict concavity.

6.3. **Role of the Constraint Qualification**

We'll now show that in the absence of the caveat about the constraint qualification, the KKT theorem above would be false.

- the constraint qualification relates to the mysterious "bizarre exception" that is added as a caveat to the mantra.

- In Fig. 3 below, the gradient vectors for the constraints satisfied with equality are not linearly independent, and don't span the domain of the objective function, so that you can't write the objective function as a linear combination of them. This doesn't mean that we don't have a maximum at the tip of the constraint set.
- A sufficient condition for the CQ to be satisfied is that the gradients of the constraints satisfied with equality at $\bar{\mathbf{x}}$ form a linearly independent set.
- Another way of stating the above is as follows: let $M(\mathbf{x})$ denote what's left of $Jg(\mathbf{x})$ (i.e., the Jacobian matrix of $\mathbf{g}$ evaluated at $\mathbf{x}$), *after you've eliminated all rows of this matrix corresponding to constraints satisfied with strict inequality at* $\mathbf{x}$; a sufficient condition for the CQ to be satisfied at $\mathbf{x}$ is that the matrix $M(\mathbf{x})$ has full rank.
- If the constraint qualification isn't satisfied then you could have a maximum at $\bar{\mathbf{x}}$ but wouldn't know about it by looking at the KKT conditions. at this point.
- The right way to think about what goes wrong is that when the CQ is violated, the information conveyed by the gradient vectors doesn't accurately describe the constraint set.
- Saying this more precisely, when we use the KT conditions to solve an NPP, what we are actually doing is solving the *linearized version* of the problem. By this I mean, we replace the boundary of the constraint set, as defined by the level sets of the constraint functions, by the piecewise linear surface defined by the *tangent planes* to the level sets. In order for the KT conditions to do their job, it had better be the case that in a neighborhood of the solution you are considering, this "linear shape" looks, *locally*, pretty much like the shape of the true nonlinear constraint set for the problem.
- In the case of our example, the tangent lines to the two level sets, at the tip of the pointy cone, are parallel to each other, so that the "linearized problem" is a line that goes on for ever.
  - Given the objective function that we've drawn, the linearized problem has *no* solution, you just go on and on forever along the line.
  - the constraint set for the true problem doesn't look anything like a line, so that the linearized problem is completely misleading

- observe that if you "bumped" the original problem a little bit, so that the tangent planes at the tip were not parallel to each other, then the close relationship between the original and the linearized problmes would be restored.
- what determines whether the linearized problem accurately represents the original problem? Answer is apparent from comparing the relatinoship between the gradient vectors to the tangent planes for the original, pathological problem to this relationship for the bumped, well-behaved problem: in the former case, the two gradient vectors are collinear.
- More generally, the linearized problem will *locally* accurately represent the original nonlinear problem at a point if the gradient vectors of the constraints that are satisfied with equality at that point form a linear independent set.

It follows from the above that there are two special cases in which the constraint qualification is satisfied vacuously.

(1) if all of the constraints are linear, (more precisely, if all of the constraint functions have affine level sets). In this case, trivially, the linearized problem will be identical to the original problem.
(2) if there is only one constraint, whose gradient is non-zero

It's important to keep in mind that the constraint qualification is a *sufficient* condition that ensures that the KT conditions are *necessary* for a solution. Some people have wondered the following: "could I find a point that satisfies the KT conditions, leading me to believe that I've found a solution to the NPP, but then the constraint qualification turns out to be violated, so that candidate solution turns out not to be a solution?" The answer to this question is a resounding NO. If the constraint qualification is violated, there may be solutions to the NPP that *don't* satisfy the KT conditions. On the other hand, the KT conditions may lead indeed me to believe that I've found a solution to the NPP, which for reasons *other than* the constraint qualification failing, is not in fact a solution.

If you say that finding a point that satisfies the KT conditions is a "positive" and not finding such a point is a "negative", then we can talk about "false positives" (in the sense of a positive result of a test, even though the thing you're testing for (say pregnancy) hasn't happened) and "false negatives" (the test says you aren't pregnant when you are).

In KT theory, we've seen both kinds of "falseness": there are several ways that we can get "false positives" in the sense that the KT conditions are satisfied at some point which *does not* solve the NPP. We'll see some of these in a minute. However, provided your function is differentiable and the constraint set is compact, the only way to get a "false negative" is for the CQ to be violated (e.g., Varian's "Hershey Kiss" example (Fig. 3 below)). Thus, provided the CQ is not violated, there will be no "false negatives" and the maximum (if it exists) must satisfy KT.

6.4. **Demonstration that KKT conditions are necessary**

Rather than prove the result generally, we are just going to look at special cases. Throughout this subsection, we are going to ignore the question of whether the KTT conditions are sufficient for a solution. The pictures we are going to consider are of functions so well behaved that if the KTT conditions are satisfied at a point then that point is a solution.

The case of one inequality constraint: With one constraint, the picture is easy: the condition is that $\bigtriangledown f(\bar{\mathbf{x}}) = \lambda \bigtriangledown g(\bar{\mathbf{x}})$, for some nonnegative scalar λ. Also, with one constraint the CQ is vacuous. In Fig. 4

- The constraint is the condition that $g(\mathbf{x}) \leq b$. That is, the only $\mathbf{x}$'s we are allowed to consider are $\mathbf{x}$'s for which the function g assigns values weakly less than b.
- Three candidates in the picture $\mathbf{x}^1$, $\mathbf{x}^2$ and $\mathbf{x}^3$. Each curve represents a level set of some function thru one of these points. Note: each level set belongs to a *different* function!
- Point $\mathbf{x}^2$ isn't a maximum; upper contour set intersects with constraint set; Point $\mathbf{x}^3$ isn't either. Point $\mathbf{x}^1$ is the only one with the property that there's nothing in the *strict* upper contour set that is also in the constraint set.

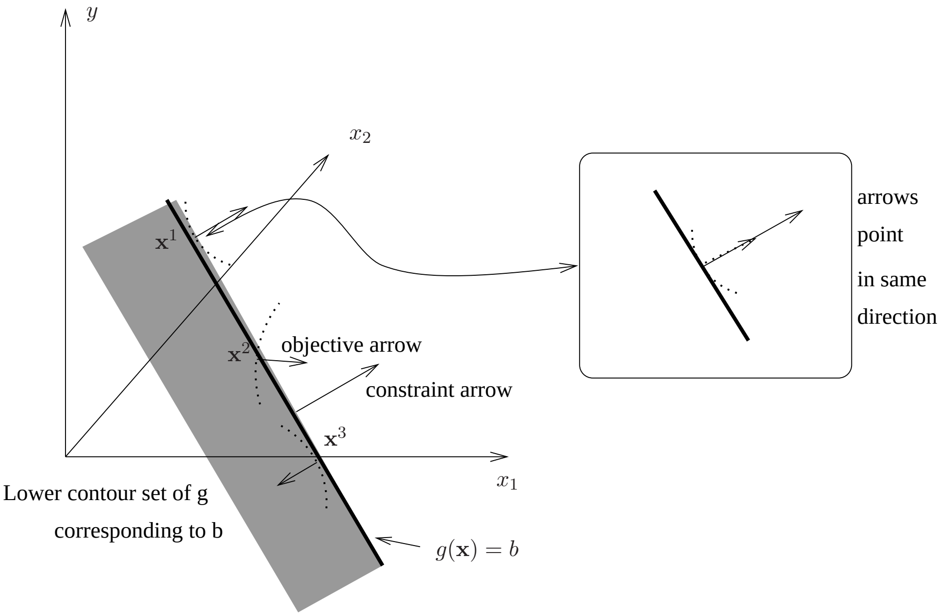

Figure 4. Constrained max problem with one inequality constraint

- Now put in the arrows. Call the first the *objective arrow* (i.e., the gradient of the objective function) and the second the *constraint arrow* (i.e., the gradient of the constraint function). Which way do the objective arrows point?
  - arrow points in direction of increase
  - this must be the upper contour set
  - upper contour set is convex
  - Conclude: we know that the arrow for a level set points into the corresponding *upper contour* set, which for a quasi-concave function is the *convex* set bounded by the set.
- Which way does the constraint arrow point? Must be NE because the lower contour set is SW.
- Now characterize the condition for a maximum in terms of the arrows. Answer is that arrows lie on top of each other and point in the same direction (as opposed to lying on top of each other but point at 180 degrees to each other).
- Conclude in terms of the mantra: arrow lies in the positive cone defined by the unique binding constraint.

Question: Suppose that for each of the level sets of f, the gradient vector at every point on this level set points SW as in point $\mathbf{x}^3$? What can you conclude about the solution to the maximization problem?

Answer: There is no solution. Certainly there can't be a solution in which the constraint is binding. And there can't be an unconstrained solution either. You can *always* move a bit further southwest and increase the value of f

More formally, for the case of one inequality constraint, the KKT condition is that there is a *nonnegative collinear* relationship between the gradients of the objective and the constraint functions. To see that this condition is necessary, suppose you don't have nonnegative collinearity, i.e., $\bigtriangledown f(\bar{\mathbf{x}})$ is not a nonnegative scalar multiple of $\bigtriangledown g(\bar{\mathbf{x}})$. Then you can find a vector $\mathbf{dx}$ such that $\bigtriangledown g(\bar{\mathbf{x}}) \cdot \mathbf{dx} < 0$ and $\bigtriangledown f(\bar{\mathbf{x}}) \cdot \mathbf{dx} > 0$. (Takes a bit of linear algebra work to prove this but it's obvious diagramatically.) But this means that $\mathbf{x} + \mathbf{dx}$ satisfies the inequality constraint and (by Taylor's theorem) increases f provided $\mathbf{dx}$ is sufficiently small.

The case of one equality constraint, $g(\mathbf{x}) = b$:

- Could write the problem as

$$\max f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq b; g(\mathbf{x}) \geq b.$$

  i.e., solution must be in the lower contour set of g corresponding to b; it must also be in the upper contour set of g corresponding to b; intersection of these two constraints is the level set of g corresponding to b.
- Trouble is that this formulation of the problem doesn't fit our general specification. We are required to specify the constraint set as the intersection of LOWER contour sets.
- Hence rewrite $g(\mathbf{x}) \geq b$ as $-g(\mathbf{x}) \leq -b$. That is, our problem is

$$\max f(\mathbf{x}) \text{ subject to } g(\mathbf{x}) \leq b; -g(\mathbf{x}) \leq -b.$$

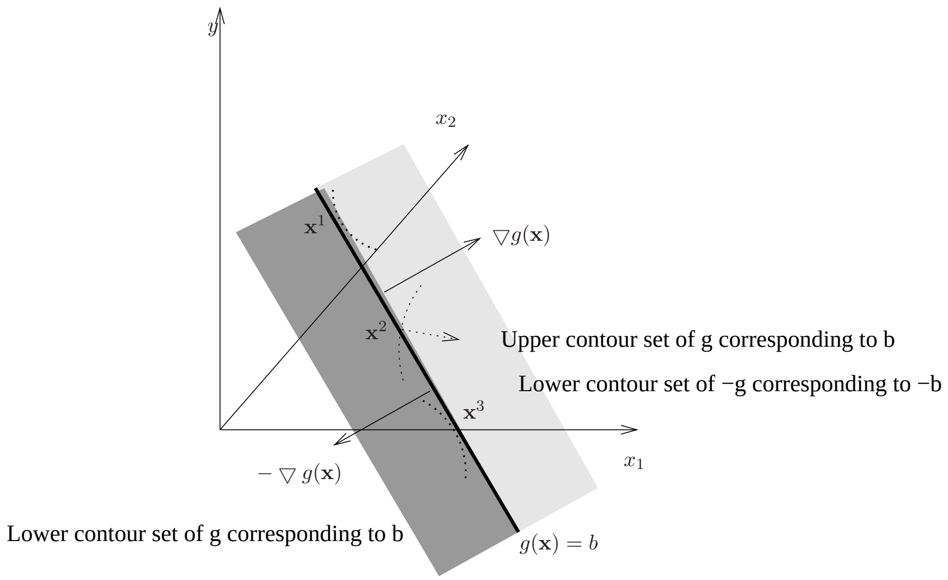

FIGURE 5. One equality constraint viewed as two inequality constraints

What can we say about the necessary conditions for a maximum in terms of the gradient vector of the objective function and the gradient vector(s) of the *binding* constraint(s). Remember that whether we are talking about the level set passing through $\mathbf{x}^1$ or the one passing through $\mathbf{x}^3$, the objective function is quasiconcave, and this implies something about the direction in which the arrow is pointing. Consider the two candidates, $\mathbf{x}^1$ and $\mathbf{x}^3$:

- For point $\mathbf{x}^1$, the gradient of f is pointing NE. Points NE of the constraint at $\mathbf{x}^1$ are assigned HIGHER values by g; just as in the previous case, the constraint $g(\mathbf{x}) \leq b$ is binding.
- For point $\mathbf{x}^3$, the gradient of f is pointing SW. Points SW of the constraint at $\mathbf{x}^1$ (and therefore at $\mathbf{x}^3$) are assigned LOWER values by g. Hence the original constraint isn't binding (i.e. $g(\mathbf{x}) \leq b$): if all you had to worry about was this constraint, you'd keep moving SE. BUT, the second constraint *is* binding. Moving SW increases the value of the function $-g$ and this is inadmissible.

Note the distinction here between a constraint that is binding and one that is satisfied with equality. In this case, both constraints, $g(\mathbf{x}) \leq b$ and $g(\mathbf{x}) \geq b$, are satisfied with equality. Only one has the property that if you took the constraint away, the maximized value of the objective function would increase.

- In either case, what can we say about the objective arrow and the arrow corresponding to the binding constraint? Ans: Pointing in the same direction. In other words, the gradient of the objective function belongs to the positive cone defined by the gradient(s) of the binding constraint(s).

Conclude that when the problem is viewed as an inequality constrained problem, I can say that a necessary condition for a maximum is that the arrow for f is in the *nonnegative* cone defined by the constraints satisfied with equality. It is also in the positive cone defined by the gradient of the binding constraint. The mantra lives!

Since we know that f is quasi-concave, and that $g(\cdot)$ is linear, then the above condition is both necessary and sufficient.

In this case, let $g^1 = g$ and $g^2 = -g$. The KT conditions say that

$$\bigtriangledown f(\bar{\mathbf{x}}) \quad = \quad \lambda_1 \bigtriangledown g^1(\bar{\mathbf{x}}) + \bigtriangledown \lambda_2 g^2(\bar{\mathbf{x}}), \text{ for some pair } (\lambda_1, \lambda_2) \geq 0. \tag{1}$$

For the purposes of computation, however, it is convenient to recognize that the above condition is equivalent to: $\bigtriangledown f(\bar{\mathbf{x}}) = \lambda \bigtriangledown g(\bar{\mathbf{x}})$, for some *arbitrary* scalar λ, i.e., (i.e., λ in this case is *not* restricted to be nonnegative). To see why the two conditions are equivalent note that

(1) $\lambda_1 \bigtriangledown g^1(\bar{\mathbf{x}}) + \lambda_2 \bigtriangledown g^2(\bar{\mathbf{x}}) \quad = \quad (\lambda_1 - \lambda_2) \bigtriangledown g(\bar{\mathbf{x}})$

(2) At most one of the two λ's can be non-zero.
    (a) if $(\lambda_1 - \lambda_2) > 0$ then the g^2 constraint *cannot* be binding, therefore λ_2 must be zero.
    (b) if $(\lambda_1 - \lambda_2) < 0$ then the g^1 constraint *cannot* be binding, therefore λ_1 must be zero.
    (c) if $(\lambda_1 - \lambda_2) = 0$ then $\bigtriangledown f$ must be zero, in which case *neither* constraint can be binding, so that $\lambda_1 = \lambda_2 = 0$.

    Because of this (i.e., because at most one λ can be non-zero), we can collapse the two λ's into one, as follows:

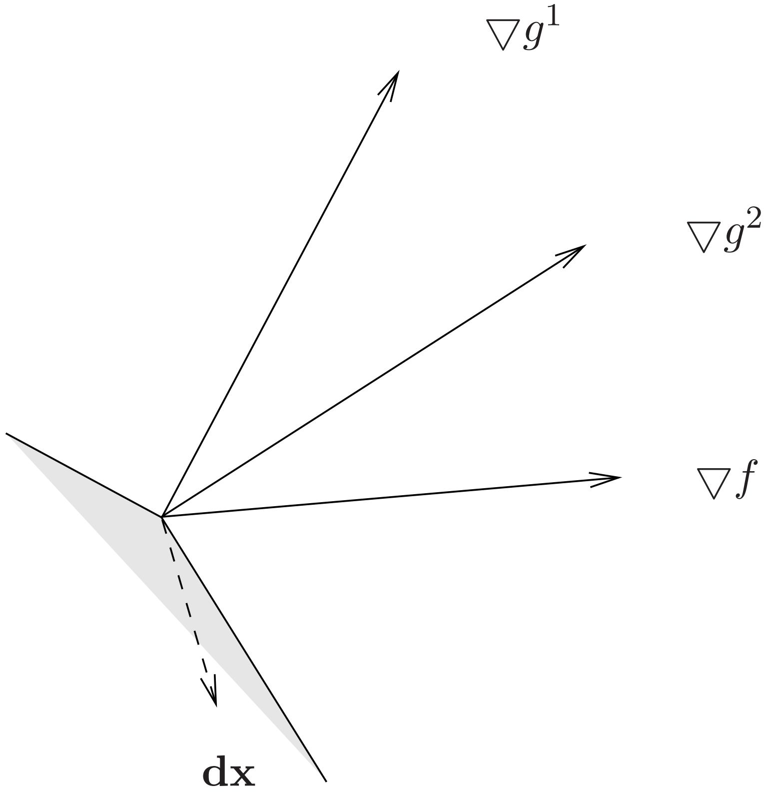

FIGURE 6. Graphical Illustration of the KKT conditions: two constraints

(3) Define $\lambda = \begin{cases} \lambda_1 & \text{if } g^1 \text{ is binding} \\ -\lambda_2 & \text{if } g^2 \text{ is binding} \\ 0 & \text{if neither are binding} \end{cases}$ and observe that (1) is now equivalent to

$$\bigtriangledown f(\bar{\mathbf{x}}) \quad = \quad \lambda \bigtriangledown g(\bar{\mathbf{x}}), \qquad \lambda \in \mathbb{R} \tag{1'}$$

<u>The two constraint case</u>: With two constraints the condition says that at the solution the gradient vector $\bigtriangledown f(\bar{\mathbf{x}})$, can be written as a nonnegative linear combination of the gradients of the constraints that are satisfied with equality

What does this mean geometrically:

- means that the gradient vector of the objective function points into the cone defined by the gradients of the constraints that are satisfied with equality. Go over what it means for a vector to be inside the positive cone defined by two others. Show geometrically that if a vector $\mathbf{x}$ is inside the cone, you can take positive scalar multiples of the vectors that define the cone, and reconstruct $\mathbf{x}$. Otherwise, you can't reconstruct $\mathbf{x}$ with positive coefficients.

- to see why must this condition hold geometrically, we draw a picture (Fig. 6) that has *only* the gradients drawn in (none of the constraint sets or indifference curves will be drawn in) and argue from the vectors *alone* why the above condition is necessary.
- Suppose the above condition is violated, so that $\bigtriangledown f(\bar{\mathbf{x}})$ lies outside the cone defined by the gradients of the constraints that are satisfied with equality. Clearly, we can then draw a line (the horizontal dotted line in Fig. 6) such that the gradient vectors of all of the constraints lie on one side of the line, and the gradient vector of the objective function lies on the other side. (Again this is obvious geometrically, but requires some work to prove rigorously.)
- We can now draw a vector $\mathbf{dx}$ that makes an acute angle with $\bigtriangledown f(\bar{\mathbf{x}})$ but an obtuse angle with all of the constraint vectors. We can *always* do this: make $\mathbf{dx}$ arbitrarily close to the line perpendicular to the first dotted line, which ensures that $\mathbf{dx}$ will make an obtuse angle with the constraint gradients, but must make an acute angle with $\bigtriangledown f(\bar{\mathbf{x}})$ since both vectors are trapped within the quadrant defined by the dotted lines.
- Observe that it makes an obtuse angle with all of the constraint vectors.
- Reason from there:
  - add $\mathbf{dx}$ to $\bar{\mathbf{x}}$
  - observe that $f(\bar{\mathbf{x}} + \mathbf{dx}) - f(\bar{\mathbf{x}}) \approx \bigtriangledown f(\bar{\mathbf{x}}) \cdot \mathbf{dx} > 0$;
  - i.e., moving in this direction increases the objective function.
  - similarly, for all j such that the j'th constraint is satisfied with equality, observe that $g^j(\bar{\mathbf{x}} + \mathbf{dx}) - g^j(\bar{\mathbf{x}}) \approx \bigtriangledown g^j(\bar{\mathbf{x}}) \cdot \mathbf{dx} < 0$, i.e., reduces the value of all of the constraints that are satisfied with equality.
  - for j's that aren't, if $\mathbf{dx}$ is sufficiently small, then you can increase the value of $g^j(\cdot)$ a little bit and still be less than b^j.
  - in other words, you can move a little in the direction of $\mathbf{dx}$ and stay within the constraint set, yet increase the value of f.
  - This establishes that $\bar{\mathbf{x}}$ couldn't have been a maximum.